

Synthese
<https://doi.org/10.1007/s11229-019-02091-2>



Safety, domination, and differential support

Charles Neil¹ 

Received: 3 March 2018 / Accepted: 9 January 2019
© The Author(s) 2019

Abstract

In a recent paper “Safety, Sensitivity, and Differential Support” (*Synthese*, December 2017), Jose Zalabardo argues that (contra Sosa in *Philos Perspect* 33(13):141–153, 1999) sensitivity can be differentially supported as the correct requirement for propositional knowledge. Zalabardo argues that safety fails to *dominate* sensitivity; specifically: some cases of knowledge failure can only be explained by sensitivity. In this paper, I resist Zalabardo’s conclusion that *domination failure* confers differential support for sensitivity. Specifically, I argue that counterexamples to sensitivity undermine differential support for sensitivity. Using Zalabardo’s modal framework, I consider a less demanding modal condition, what I call *weak sensitivity*, and I explain how weak sensitivity avoids an influential counterexample to sensitivity. However, I argue that we can subvert that counterexample only by abandoning Zalabardo’s case for *domination failure*. So either way, we cannot differentially support sensitivity.

Keywords Safety · Sensitivity · Modal epistemology · Differential support · Epistemic luck

The paper is structured as follows:

- §1. I foreground standard modal versions of the safety and sensitivity conditions. I retrace Zalabardo’s argument that sensitivity entails safety.
- §2. I argue that *domination failure* alone doesn’t confer differential support for sensitivity; counterexamples to sensitivity affect differential support.
- §3. I distinguish between *weak sensitivity* and *strong sensitivity*. I argue that safety fails to dominate sensitivity only if we endorse *strong sensitivity*, which succumbs to counterexamples which undermine differential support for sensitivity.

I am grateful to two referees for this journal for their helpful comments.

✉ Charles Neil
neil.charles.13@ucl.ac.uk

¹ Department of Philosophy, University College London, London, UK

1 Domination and translation keys

Consider two venerable conditions for (propositional) knowledge, the sensitivity and safety conditions.¹ Let us say that a subject *S* has a *sensitive* belief in some proposition *p* just in case:

“were it not the case that *p*, then *S* would not believe that *p*”²

Formally:

Sensitivity: $\sim p \Rightarrow \sim Bp$

(Throughout this paper, let “ \Rightarrow ” denote the subjunctive conditional and let “ \rightarrow ” denote the material conditional).

Nozick (1981) regards sensitivity as a necessary condition for knowledge. In contrast, Sosa (1999) argues that sensitivity should be replaced with the following subjunctive condition, which he calls *safety*:

“*S* would believe that *p* only if it were so that *p*”³

Formally, Sosa writes safety as the contrapositive of sensitivity, as follows:

Safety: $Bp \Rightarrow p$

As Sosa reminds us, sensitivity and safety are not equivalent because subjunctive conditionals do not contrapose.⁴

Sosa argues that sensitivity cannot be differentially supported as the correct requirement for propositional knowledge. He does so in defence of neo-Mooreanism, the view that we can and do know the negation of radical sceptical hypotheses.⁵

Zalabardo summarises Sosa’s argument for *the denial of differential support for sensitivity* as the claim that “every virtue that can be claimed for sensitivity as a necessary condition for knowledge would be matched by conferring this status on safety instead” (p. 2). Zalabardo argues that the denial of differential support requires the truth of two premises, the conjunction of which he calls *domination*:

¹ Nozick also relativizes the sensitivity conditional to the method of belief formation “*M*” which *S* employs in the actual world to believe that *p* (p. 179). Similarly, safety has been relativized to the method or basis which one uses in the actual world (See Sosa 2007). In this paper, I will be concerned only with the un-relativized versions of sensitivity and safety, which are the focus of Zalabardo’s paper.

² On an interpretive note, Nozick intended “sensitivity” to refer to the conjunction of (what I’ve called) the sensitivity condition *and* an adherence condition, namely $[p \Rightarrow Bp]$ (Nozick, p. 176). I will not be discussing adherence here. I shall use sensitivity to refer only to the subjunctive $[\sim p \Rightarrow \sim Bp]$.

³ Sosa (1999, p. 142).

⁴ Sosa (1999, pp. 149–150).

⁵ Sensitivity, but not safety, is inconsistent with neo-Mooreanism; the view that we can and do know that radical sceptical hypotheses don’t obtain (in the spirit of G.E. Moore’s pronouncements (1918, 1925, 1939)). Take, for example, the hypothesis that one is a disembodied brain-in-a-vat (BIV). Notice that one’s belief that \sim BIV is safe but insensitive. For *not easily* would one falsely believe that \sim BIV (hence safety is satisfied), but *if* BIV were true *then* one would falsely believe that \sim BIV (hence sensitivity is violated). By replacing sensitivity with safety, Sosa blocks the argument against neo-Mooreanism based on the contention that one’s belief that \sim BIV is insensitive.

(D1) If a belief that has the status of knowledge is sensitive, then it is also safe.

(D2) If a true belief that doesn't have the status of knowledge is insensitive, then it is also unsafe.⁶ (p. 2)

More schematically:

(D1) $[Kp \ \& \ (Bp)\text{SENSITIVE}] \rightarrow (Bp)\text{SAFE}$.

And

(D2) $[\sim Kp \ \& \ \sim (Bp)\text{SENSITIVE}] \rightarrow \sim (Bp)\text{SAFE}$.⁷

Domination fails if either **D1** or **D2** is false (for an inclusive disjunction). Zalabardo rejects *domination* by conceding **D1** and rejecting **D2**.

I will now examine Zalabardo's argument for **D1**. **D1** is true because, Zalabardo argues, sensitivity entails safety. To understand why sensitivity entails safety, let us first examine how safety requires a different *translation key* to sensitivity.

What is a *translation key* for a subjunctive conditional? A *translation key* for a subjunctive conditional is a way of unearthing the truth value of that subjunctive. On David Lewis's view, for example, the subjunctive conditional "if it were the case that p , then it would be the case that q " is *true* just in case in the closest possible worlds to the actual world " α "⁸ in which p obtains, q also obtains (Lewis 1973a, b). I will refer to this as the *standard translation key* (ST). More schematically:

ST: $p \Rightarrow q$ is true just in case in all the worlds in which p is true that are at the shortest distance from α , q is also true.

Sensitivity theorists embrace **ST**. They say that S 's true belief that p is sensitive just in case the *closest worlds* to α in which the antecedent is true (i.e. in which $\sim p$ holds) are worlds in which the consequent is true (i.e. in which $\sim B(p)$ holds).⁹ However, if safety theorists adopt **ST**, they're in trouble. Their problem is this: if S has a true belief that p in the actual world α , then the closest possible world to α in which S truly believes that p *just is* α . The closest world to $\alpha = \alpha$, and therein lies the problem with **ST** for subjunctive conditionals with true antecedents (such as safety). Adopting **ST** for safety renders all true beliefs trivially safe.

Safety theorists, as Zalabardo observes, must use a different translation key to avoid triviality. Specifically, safety is usually made a function of what one believes at "close" possible worlds which lie beyond α . Let us call this the *non-standard translation key* (**NST**). Let us formulate this key as follows, where " d " denotes a sphere of *close* possible worlds:

NST: $p \Rightarrow q$ is true just in case in every world in which p is true that is at a distance of d or less from α , q is true.

⁶ Zalabardo (p. 2) qualifies the consequent: non-modal shortcomings may account for why a sensitive belief fails to constitute knowledge.

⁷ D1 holds that *if* one knows that p (Kp) AND one's belief that p (Bp) is sensitive, *then* Bp is safe. D2 holds that *if* one fails to know that p ($\sim Kp$) AND Bp is insensitive, *then* Bp is unsafe.

⁸ Throughout this paper, let " α " denote actuality.

⁹ See Zalabardo (2017, p. 3).

Zalabardo observes that by using **NST**, safety becomes the following non-trivial condition:

Safety (NST): S's belief that p is safe just in case, in every world in which S believes that p that is at a distance of d or less from α , p is true. (Zalabardo, p. 3)

At this point, Zalabardo argues that if we adopt **NST** for safety, then we should adopt **NST** for sensitivity too. Zalabardo reasons as follows:

it seems wrong to make the truth conditions of a subjunctive depend on the truth value of its antecedent. Take, for example, the subjunctive “if the currency were devalued, interest rates would go up.” On the current proposal, if the currency is actually devalued, the truth of the subjunctive will require interest-rate rises, not only in actuality, but also in the range of conditions obtaining in all the worlds at a distance of d or less in which the currency is devalued, e.g. at a range of levels of taxation, inflation, etc., but if the currency is not devalued, it would suffice for the truth of the subjunctive that interest rates go up in the specific circumstances obtaining at the nearest worlds in which the currency is devalued, e.g. with the precise levels of taxation, inflation, etc. present in those worlds. (p. 5)

Zalabardo concludes:

I find this counterintuitive. If, in order to avoid making safety redundant, we look at a wide range of worlds to determine the truth value of the subjunctive, we need to apply the same approach to sensitivity. (p. 5)

Zalabardo argues that if we employ **NST** for sensitivity, then sensitivity entails safety. To understand why, let us reproduce Zalabardo's argument. Let $CT(p)$ denote the distance from α of the closest possible world to α in which p is true. Notice that $CT(p) = 0$ for whenever p is true. Therefore, for actually obtaining beliefs, **NST** implies that a belief that p is safe just in case p is true in all worlds in which one believes that p up to d (p. 5). Conversely, adopting **NST** for sensitivity now yields the following condition:

Sensitivity (NST): S's belief that p is sensitive just in case: in every world in which p is false that is at a distance of $CT(\sim p) + d$ or less from α , S doesn't believe p . (Zalabardo, p. 5)¹⁰

On this construal, sensitivity entails safety. To see why, let us follow Zalabardo and understand safety and sensitivity as properties which exclude *worlds in which one falsely believes that p* —what Zalabardo terms *error worlds*. The only difference between safety and sensitivity, Zalabardo contends, is that safety but not sensitivity imposes a *fixed lower bound* on which error worlds knowledge excludes (the bound is fixed for safety because $CT(p) = 0$ for subjunctives with true antecedents). Safety prohibits the existence of error worlds up to distance d only. By contrast, sensitivity demands that at the closest $\sim p$ worlds (however distant!) one doesn't believe that p

¹⁰ Nozick employs NST for his adherence condition ($p \Rightarrow Bp$). For the subjunctive is trivially true for all true beliefs unless one considers whether Bp obtains at *close* worlds in which p obtains. Nozick claims that ($p \Rightarrow Bp$) holds true if “not only does he actually truly believe that p , but in the “close” worlds where p is true, he also believes it” (p. 176).

and one doesn't falsely believe that p up to d . Crucially, sensitivity will always require the elimination of error worlds up to d . Notice that, on this view, a belief that p will be safe but insensitive just in case the closest $\sim p$ worlds fall outside the d sphere and if in one of those $\sim p$ worlds one falsely believes that p . In this respect, sensitivity is the more demanding condition and **D1** is true.

Sensitivity and safety are manifestly *anti-luck conditions*. A widely-held intuition is that knowledge is incompatible with certain kinds of epistemic luck. Consider the familiar stopped-clock example.¹¹ Imagine that S forms a true belief that *the time is 12.00* (henceforth " $t12$ "), but where unbeknownst to S the clock stopped exactly 24 h beforehand. Intuitively S doesn't know that $t12$ because her true belief seems luckily true. Both sensitivity and safety accommodate this verdict. After all, S 's belief is *insensitive* because the closest worlds in which $\sim t12$ (for example, the worlds where the clock stops at 11.59 or 12.01) are worlds in which S falsely believes that $t12$, and S 's belief is *unsafe* due to the many close worlds in which S falsely believes that $t12$. Let us call the species of knowledge-excluding epistemic luck which arises in such cases, *veritic (epistemic) luck* (following Engel 1992).¹² Modalized sensitivity and safety eliminate a wide class of cases of veritic luck; this includes Gettier-type cases and true beliefs which result from wishful thinking or guesswork. Furthermore, for a wide class of cases, sensitivity and safety are on par in countering veritic luck (Pritchard 2015).

Importantly, both sensitivity and safety eliminate veritic luck without committing one to a thoroughgoing scepticism about knowledge. For example, take your putative knowledge that H [I have hands]. Notice that, to the extent that you know that H , your belief that H is both *sensitive and safe*. Modally speaking, the closest $\sim H$ worlds are ones in which you don't falsely believe that H (for example, the worlds in which you're an amputee), and so your belief is sensitive. Similarly, H is true in most close worlds in which you believe that H , and so your belief is safe. Sensitivity and safety corroborate many cases of knowledge possession because subjunctive conditionals differ from logical entailments. For the sensitivity subjunctive, we need only examine S 's doxastic dispositions at the *closest* $\sim H$ worlds. And for safety, we need only examine whether H holds at the *close* worlds in which S believes that H . Both proposals are thus compatible with the existence of radical error worlds, such as the world in which you're a disembodied brain-in-a-vat (BIV) and in which you falsely believe that H . However, BIV worlds are not the *closest* $\sim H$ worlds (hence sensitivity is satisfied), nor should they be included in the domain of *close* worlds in which you believe that H (hence safety is satisfied). The existence of radical error worlds doesn't render your belief that H insensitive or unsafe.

¹¹ See Russell (1948, p. 170) for original example. For further discussion of the stopped-clock example in the context of safety and sensitivity, see Pritchard (2008). For an explanation of how true beliefs may be "Gettiered", see Zagzebski (1994).

¹² Engel (1992) defines a belief as being "veritically lucky" when it is a matter of luck, given one's evidential situation, that one's belief is true (p. 67). Note that anti-luck conditions needn't require the elimination of epistemic luck *simpliciter*. For example, Engel distinguishes "veritic luck" from "evidential luck", where the latter pertains to the luck that one is in an evidential situation E , which is to be distinguished from the luck which pertains to the truth of the belief given that one is in E (ibid). Engel argues that evidential luck is compatible with knowledge. See Unger (1968) for an analogous distinction.

Both sensitivity and safety exclude (a wide class of) cases of veritic luck from the extension of knowledge, while also retaining much of our putative knowledge. But despite both conditions yielding identical predictions for a wide range of cases, I will argue that safety rather than sensitivity is the correct anti-veritic luck condition.

2 Domination and differential support

Domination requires the truth of the conjunction [D1 & D2]. Zalabardo denies the truth of the conjunction by rejecting the second conjunct, D2.

But the first thing to notice is that \sim [D1&D2] *alone* cannot surely confer *differential support* for sensitivity. For \sim [D1&D2] is consistent with counterexamples to sensitivity. Sosa argued for *domination*, but he also propounded counterexamples to sensitivity. Sosa *replaced* sensitivity with safety because safety does not (ostensibly) succumb to counterexamples. On first blush: S knows that p only if S has a safe belief that p. Formally,

$$Kp \rightarrow (Bp \Rightarrow P)$$

But that schema tells only half the story. Because, so far, we have not eliminated the necessity (“ \Box ”) of the sensitivity condition. Sosa argues that knowledge doesn’t require sensitivity (i.e. $[Kp \rightarrow (\sim p \Rightarrow \sim Bp)]$). Thus, Sosa’s safety premise is better represented as:

$$Kp \rightarrow (Bp \Rightarrow P) \& (\sim \Box (\sim p \Rightarrow \sim Bp))$$

The above principle allows insensitive but safe beliefs to constitute knowledge. In sum: a belief having the property of insensitivity doesn’t affect its epistemic standing. It is therefore the conjunction of *domination* and Sosa’s *counterexamples* which are used to *differentially support safety* as the correct requirement for knowledge. That is, Sosa argues that there are no virtues which are exclusive to sensitivity *AND* there are some virtues which are exclusive to safety. Sosa’s best-known counterexample to sensitivity is this:

CHUTE: “On the way to the elevator I release a trash bag down the chute of my high rise condo. Presumably I know my bag will soon be in the basement. But what if, having been released, it still (incredibly) were not to arrive there? That presumably would be because it had been snagged somehow in the chute on the way down (an incredibly rare occurrence), or some such happenstance. But none such could affect my predictive belief as I release it, so I would still predict that the bag would soon arrive in the basement. My belief seems not to be sensitive, therefore, but constitutes knowledge anyhow, and can correctly be said to do so.” (Sosa 1999, p. 145)

Does safety accommodate CHUTE? Safety theorists will need the following verdict: if the chute is a regular smooth chute, then the bag-snagging world is *not close*. Conversely, if the chute has structural irregularities which leave it prone to bag-snagging,

then the bag-snagging world is *close*. Notice that if the chute is prone to bag-snagging, then the example may also fail to elicit the intuition that the subject *knows*. If this is the case, then our safety intuition in CHUTE is covariant with the epistemic intuition that the subject knows the target proposition.¹³

Nevertheless, even if the chute is smooth, the closest error worlds in CHUTE are not nearly as distant as the error worlds in which you or I falsely believe that we're not a disembodied brain-in-a-vat, for example. That is, on a similarity ordering of possible worlds, the error worlds in CHUTE are not *radically dissimilar* to the actual world. After all, not *that much* would have to change—modally speaking—for the bag to snag. The error world is not distant enough for it to be uncontroversial that safety is satisfied. The worry, therefore, is that there is no principled way of excluding bag-snagging worlds from the domain of *close* worlds. The case rests on our intuitions about whether the error worlds are “close”, and here—I suspect—the case may fail to elicit clear reactions in either direction.

I want to concede that safety as Zalabardo formulates it fails to straightforwardly accommodate CHUTE (This concession is benign since I shall argue that sensitivity can't be differentially supported). If this is so, then there are construals of safety which *do* accommodate CHUTE. Note first that Zalabardo construes safety *strongly*. For example, Zalabardo argues that S's belief that *p* is safe just in case in *every* world in which S believes that *p* that is at a distance of *d* or less from actuality, *p* is true (p. 3). The inclusion of the universal quantifier “all” makes this safety condition *modally strong* because that quantifier prohibits the existence of *any* error world up to distance *d* or less from α . If CHUTE is a counterexample to *sensitivity*, then—so it goes—CHUTE is a counterexample to *strong safety* (for there exists at least *one* close possible bag-snagging world in which one falsely believes that the bag is in the basement). Hence, it has been claimed, CHUTE doesn't motivate strong safety over sensitivity. However, we can see straight away that by adopting the non-standard translation key (NST) safety theorists needn't endorse strong safety. Consider, for example, a more liberal version of safety:

Weak safety: S's belief that *p* is weakly safe just in case in *most* worlds in which S believes that *p* that are at a distance of *d* or less from α , *p* is true.¹⁴

CHUTE is not a counterexample to weak safety, because in *most* close possible worlds, one's belief in the target proposition is true (Pritchard 2012, 2015). Indeed, in virtually all close possible worlds in which one believes that [the bag is in the basement], one's belief in this proposition is true. This matters. Because when we assume strong safety, it seems that both sensitivity and safety conditions are roughly on par in the number of counterexamples to their necessity. However, the question of how *many worlds* are relevant to safety arises because safety theorists embrace NST. By embracing NST for safety, we can see straight away that there arises the question of how *strongly* or *weakly* safety ought to be construed. It is thus open to safety theorists to argue that *sensitivity fails to dominate* weak safety. To argue, in other words, that there are cases

¹³ My thanks to an anonymous referee for this journal on this point.

¹⁴ This is based on Greco (2012).

of knowledge possession which pass the *weak safety test*, but which fail the *sensitivity test*. In effect, to argue as follows:

(D1*) $[Kp \ \& \ (Bp) \text{ weakly safe}] \rightarrow (Bp) \text{ sensitive.}$

(D2*) $[\sim Kp \ \& \ \sim (Bp) \text{ weakly safe}] \rightarrow \sim (Bp) \text{ sensitive.}$

It cannot be said that sensitivity reaps every advantage reaped by weak safety. If sensitivity entails weak safety, while that guarantees D1, entailment doesn't guarantee D1*. Entailment doesn't stop sensitivity failing to dominate weak safety. And if sensitivity doesn't dominate weak safety, then we can't *differentially support* sensitivity.

Note, however, that there are cases of knowledge failure which are explained by strong safety but not by weak safety. Consider, for example, the following LOTTERY case. Imagine one buys a lottery ticket where the odds of winning are 60 million to 1. Despite these long odds, one's pessimistic belief that L [I am in possession of a losing ticket] intuitively falls short of knowledge before the result is announced. Nevertheless, in *most* close worlds one's belief that L is true, and so one's belief that L is weakly safe. In contrast, so it goes, there exists at least one close world in which one has a false belief that L , hence one's belief that L violates strong safety.¹⁵ Strong safety succeeds here where weak safety fails, such that the retreat from strong to weak safety may seem ad-hoc.

Two points here. First, it is unclear that safety theorists are wedded to a construal of safety which can't accommodate both CHUTE and LOTTERY. Consider Duncan Pritchard's defence of safety (2008, 2012, 2015). Pritchard construes safety as an anti-veritic luck condition, one which explains cases of knowledge failure such as LOTTERY and Gettier-type cases. Importantly, Pritchard understands luck as a modal notion which admits of degrees, where the degree of luck pertaining to an event is proportionate to the modal proximity of worlds in which that same event fails to obtain. Pritchard's safety condition is motivated by an extensive modal account of luck (see Pritchard 2015). Let's focus on the account's implications.

What is arguably problematic about LOTTERY is the very close modal proximity of the worlds in which one falsely believes that L . After all, as Pritchard reminds us, the closeness of lottery error worlds explains the lure of the game. In this case, we need only *one* error world to render the belief veritically lucky in a way which precludes knowledge. In contrast, it is not obvious that CHUTE error worlds are *very similar* to the actual world (Indeed, it's controversial whether error worlds in CHUTE are even *close*) and hence the degree of luck at stake doesn't undermine knowledge. Safety may handle CHUTE by requiring the elimination of *all very close* error worlds, with that tolerance gradually increasing as one moves further away from the actual world (Pritchard 2015, p. 101). By the time we reach the *close* error worlds in CHUTE, safety needn't display complete intolerance of error.¹⁶ Safety, at this point, still requires that one doesn't falsely believe the target proposition at *most close worlds*. It is just

¹⁵ See Greco (2003). See also Pritchard (2008, 2012, 2015). The lottery case also militates against a probabilistic construal of modal similarity, because the high likelihood of L obtaining doesn't detract from the closeness of the $\sim L$ world.

¹⁶ For discussion of weak and strong safety, see also Becker (2007), and Greco (2003), who poses the problem for safety which Pritchard (2015) addresses.

that LOTTERY violates the safety condition in virtue of safety displaying complete intolerance towards falsity at worlds which are *very close* to the actual world. Weak safety is now supplemented with a strong safety condition, one which applies only to worlds which are very close to the actual world.¹⁷

The second point, and notwithstanding Pritchard's proposal, is that we need to judge the plausibility of weak safety overall. While weak safety doesn't explain LOTTERY, weak safety does explain a class of cases of knowledge failure (Gettier cases, lucky guesses etc.) and—being a less demanding condition than strong safety—it trivially fares no worse than strong safety at corroborating intuitive cases of knowledge possession.¹⁸ On this point, we must remember that the issue we are considering here is whether sensitivity can be supported over safety as a *necessary* condition for knowledge. CHUTE may undermine the *necessity* of strong safety, but LOTTERY—as a case of knowledge failure—undermines the *sufficiency* rather than the *necessity* of weak safety.

Weak safety is an improvement on strong safety as a candidate necessary condition for knowledge because weak safety succumbs to fewer counterexamples.¹⁹ In sum: weak safety avoids at least one of the counterexamples which Sosa levied against sensitivity. This is not to vindicate weak safety as a necessary condition for knowledge, for it too might be vulnerable to counterexamples. However, at least one prominent counterexample to sensitivity doesn't affect weak safety, such that Zalabardo's case for differentially supporting sensitivity is undermined.

3 Weak sensitivity and strong sensitivity

The distinction between strong and weak sensitivity is not raised in Zalabardo's paper. However, I want to raise it here and exploit its implications for the present dialectic. I will argue that in extensional respects weak sensitivity doesn't outperform weak safety.

We may construe Zalabardo's sensitivity condition as a strong sensitivity condition. Zalabardo's sensitivity condition is *strong* because it demands that in *all* worlds up to a distance of $CT(\sim p) + d$ or less from α , S doesn't falsely believe that p . (The inclusion of the universal quantifier makes Zalabardo's condition *modally strong*). However, as with safety, sensitivity theorists who use the non-standard translation key could in principle adopt a weaker rendition, one which—like weak safety—drops the universal quantifier:²⁰

¹⁷ This is closest to Pritchard's (2008) presentation of the issue.

¹⁸ I am not saying that weak safety explains every Gettier-type case. On this point, see Greco (2003).

¹⁹ The lottery case implies that knowledge is not co-extensive with *weakly safe belief*. But safety theorists needn't *analyse* knowledge as safe belief—neither Sosa (1999), Pritchard (2012, 2015) or Williamson (2000) argue that having a safe belief is necessary and sufficient for knowledge.

²⁰ See Greco (2012), who discusses strong restricted sensitivity and weak restricted sensitivity; both conditions materially differ to Zalabardo's condition (SS) and its weaker variant (WS), insofar as restricted sensitivity makes distant $\sim p$ worlds irrelevant to the determination of sensitivity.

WS: S's belief that p is *weakly sensitive* in α just in case: at *most worlds* in which p is false at a distance of $CT(\sim p) + d$ or less from α , S doesn't falsely believe that p .

Rather than Zalabardo's strong sensitivity condition:

SS: S's belief that p is *strongly sensitive* in α just in case: at *every world* in which p is false up to a distance of $CT(\sim p) + d$ or less from α , S doesn't falsely believe that p . (Zalabardo, p. 5)

I want to make two points in what follows. First, **WS** doesn't outperform weak safety in extensional respects. That is, both conditions include in the extension of knowledge cases which should be included (CHUTE), and both fail to exclude from the extension of knowledge cases which should be excluded (ROGER). Second, **WS** provides us with no reason for abandoning the standard counterfactual analysis for sensitivity: that **WS** yields the same predictions as weak safety is explained by the claim that **WS** just is a weak safety requirement for knowledge.

Let's see how **WS** performs in extensional respects by first examining whether **WS** can accommodate CHUTE. Let "S" denote the subject who dispatches the bag down the chute and let " p " denote the proposition that *the bag is in the basement*. In the actual world, S dispatches the bag and forms a true belief that p . Consider now a range of $\sim p$ worlds up to a distance of $CT(\sim p) + d$ and consider whether or not S believes that p in most of these $\sim p$ worlds. Note that for **WS** we needn't examine only the "closest" $\sim p$ worlds (i.e. the bag-snagging worlds)—for we are now also including "close" $\sim p$ worlds in the domain of relevant $\sim p$ worlds. The **WS** condition will be satisfied if S doesn't falsely believe that p in *most* $\sim p$ worlds up to a distance of $CT(\sim p) + d$, where d denotes the range of close worlds. I think we will see that CHUTE satisfies **WS**.

The closest $\sim p$ worlds are bag-snagging worlds. But importantly, there is still a wide class of relevant $\sim p$ worlds ($\sim p$ worlds in which S doesn't believe that p) which lie further away from actuality than bag-snagging worlds, but which are paradigmatically *close*. This is due to the wide class of close worlds in which $\sim p$ holds because S has not dispatched the bag down the chute, and in these close $\sim p$ worlds S doesn't believe that p . For example, consider the many close worlds in which rather than dispatching the bag down the chute, S attends to other mundane tasks; or consider, for instance, the relatively close worlds in which the chute is closed for maintenance. Modally speaking, these $\sim p$ worlds are not radically dissimilar to the actual world; indeed many of them are very similar to the actual world. To the extent that S knows that p in the actual world, S doesn't believe that p in a wide class of close $\sim p$ worlds. The close $\sim p$ worlds described are not the *closest* $\sim p$ worlds, but they are *close* nevertheless. Furthermore, intuitively there are more *close* $\sim p$ worlds in which S doesn't believe that p than there are the narrow band of *closest* bag-snagging worlds. Therefore, CHUTE isn't a counterexample to **WS**.²¹

²¹ At a minimum, there doesn't seem to be any obvious sense in which CHUTE *violates* **WS**. However, the claim that CHUTE isn't a counterexample to **WS** is benign in the present dialectic, since I will argue that we can support weak safety over **WS** even if **WS** accommodates CHUTE.

We should take **WS** seriously as an anti-veritic luck condition because it's not implausible that **WS** yields the same prediction as weak safety for CHUTE. It may, therefore, be supposed that what sensitivity theorists need is **WS** rather than **SS**. However, I want to argue that replacing **SS** with **WS** comes at an exorbitant price for sensitivity. To see why, let us consider the following case:

ROGER: "Consider Roger, who believes \sim BIV6 (where "BIV6" is the proposition that *MI6 secretly keeps a collection of envatted brains, artificially stimulated to produce the experiences of normal embodied beings*) but for slightly unorthodox reasons. Roger doesn't believe that brain envatment is technically impossible. In fact he believes it's a common occurrence, since he heard about Putnam's thought experiment and got the wrong end of the stick. However, he is convinced that MI6 doesn't engage in these activities, the reason is that he has a friend who tells him that he works for MI6 and is always prepared to answer his questions about this service. As it happens, Roger's friend is just a cleaner in the MI6 headquarters, with no access to any classified information, and gives random but coherent answers to Roger's questions, just to humour him. When Roger asked him if MI6 keeps any envatted brains, he assures him that they didn't. It is on these grounds that Roger believes \sim BIV6." (Zalabardo, p. 8, my parentheses)

Zalabardo argues that ROGER is a case of knowledge failure explained as such by **SS** but not by safety. ROGER violates **SS** because we can expect that at *some* of the closest BIV6 worlds, Roger's friend tells him that MI6 don't keep envatted brains, and at these worlds, Roger falsely believes \sim BIV6. Hence, Zalabardo claims that safety fails to *dominate* **SS**.

Once we substitute **SS** with **WS**, however, ROGER no longer elicits the intuition that **WS** is violated. After all, ROGER's belief is (strongly) epistemically safe because the BIV6 error worlds are no less distant than the distant worlds in which *you* or *I* are a BIV; or so that is Zalabardo's claim, and that is the claim he needs for ROGER to clearly fail the safety test.²² Once ROGER is understood in this way, as strongly safe, and *a fortiori* as weakly safe, it is no longer uncontentious that **WS** does the explanatory work which weak safety fails to do. The problem at stake here is that ROGER doesn't yield the reaction that **WS** is violated. After all, ROGER is *safe* we are told, so it is unclear why ROGER is not also weakly sensitive.

For **WS** to do explanatory work here, there would need to exist *more* closest BIV6 worlds in which Roger falsely believes that \sim BIV6 than there are *close* worlds (and worlds up to the closest BIV6 worlds) in which Roger doesn't falsely believe that \sim BIV6; this suggestion is tenuous at best. Indeed, the consideration of doxastic dis-

²² Zalabardo observes "the sheer distance from actuality of the nearest BIV worlds—beyond any plausible value we might set for *d*—means that my belief in \sim BIV—and indeed any belief in this proposition by an embodied subject in the kind of world we think we inhabit—will be safe" (p. 7). Zalabardo then argues that the closest BIV6 worlds are no closer to actuality than the closest worlds in which *I* or *you* are BIV's (p. 8). However, Zalabardo's claims on p. 8 suggest a slightly different view. I will assume however that the *closest* worlds in which *I* or *you* are a BIV are *radically distant*—this is the claim that Zalabardo needs, for otherwise it won't be clear that the belief \sim BIV6 is safe and, a fortiori, it won't be clear that Roger's belief is *safe*.

positions at the closest BIV6 worlds (beyond the close worlds in which Roger doesn't falsely believe that \sim BIV6) appears to make no difference to our modal intuitions about what Roger believes at *most* of the relevant possible worlds. Therefore, **WS** fares no better than weak safety at *explaining* ROGER, which is the very case intended to show the superiority of sensitivity over safety.²³ **WS** doesn't do the explanatory work which weak safety ostensibly fails to do for ROGER.²⁴

Is there a way of adjudicating between **WS** and weak safety, given that both yield the same predictions for these cases? Do we have a safety/sensitivity impasse? I do not think we have an impasse. The impasse presupposes that we are dealing with two different conditions. However, treating **WS** as denoting the sensitivity counterfactual $[\sim p \Rightarrow \sim Bp]$ violates the standard analysis of counterfactuals. On the standard analysis, for cases where the antecedent p of a subjunctive ($p \Rightarrow q$) is presumed to be *false* in α , the counterfactual ($p \Rightarrow q$) is true just in case the consequent q obtains at the closest worlds to α where the antecedent p obtains. On the standard analysis, the sensitivity counterfactual ($\sim p \Rightarrow \sim Bp$) is false if Bp obtains at one or more of the closest $\sim p$ worlds. **WS** deviates from the standard counterfactual analysis because **WS** permits ($\sim p \Rightarrow \sim Bp$) to be true in cases where Bp obtains at one or more of the closest $\sim p$ worlds, providing that one doesn't falsely believe that p at *most* worlds up to $CT(\sim p) + d$.²⁵ **WS** violates the standard counterfactual analysis, while at the same time supplying no reason to abandon the standard analysis.

The way to preserve the standard analysis of counterfactuals is to treat **WS** as expressing the safety subjunctive ($Bp \Rightarrow p$) rather than the sensitivity subjunctive ($\sim p \Rightarrow \sim Bp$). Moreover, since **WS** and weak safety yield the same predictions for the cases under consideration, there is no reason to think that what we are dealing with here is a *sensitivity condition* rather than a *weak safety condition*.

But does weak safety also violate the standard counterfactual analysis? One final worry is that ($Bp \Rightarrow p$) may be true according to weak safety even if ($\sim p$) obtains at one or more of the "close" (Bp) worlds, which at first glance seems inconsistent with the

²³ The important point here, with respect to our modal intuitions, is that ROGER doesn't elicit the modal intuition that **WS** is violated, and that **WS** is therefore an impoverished explanation of ROGER. **WS** fails to do the explanatory work required of it.

²⁴ Weakening the non-standard sensitivity condition also renders it unable to block the neo-Moorean argument that you can and do know that you're not a brain-a-vat. After all, your belief that \sim BIV satisfies **WS** for much the same reason that it satisfies weak safety, namely that in *all* close worlds and all worlds up to the distant BIV worlds, one's belief in this proposition turns out to be true. **WS** is no less amenable to neo-Mooreanism than weak safety.

²⁵ On this point, I am indebted to a referee for this journal. **SS** also clashes with the standard counterfactual analysis, because [$\sim Bp$ obtains at the closest $\sim p$ worlds] is insufficient for **SS** obtaining. For example, **SS** will won't obtain if at the closest $\sim p$ worlds one doesn't believe that p but if there is an error world in the d sphere. However, we may at least reconcile **SS** with the standard counterfactual analysis by taking **SS** to denote the conjunction $[\sim p \Rightarrow \sim Bp \ \& \ Bp \Rightarrow p]$. On this view, the sensitivity conjunct will be satisfied in those cases where [$\sim Bp$ obtains at the closest $\sim p$ worlds], but the conjunction will be false if there is an error world within the d sphere. Zalabardo's view that sensitivity entails safety is consistent with this. As is his conclusion that "safety cannot do by itself all the work that sensitivity can do in excluding from the extension of knowledge beliefs that shouldn't be there"(p. 9). If this is so, then the claim of this paper is that the conjunction of sensitivity and safety can't—from an extensional point of view—be differentially supported over safety by itself. Note that a conjunctive view doesn't allow us to reconcile **WS** with the standard counterfactual analysis, since **WS** violates a necessary condition for the truth of the sensitivity conjunctive—namely the condition that in the closest $\sim p$ worlds, one doesn't believe that p .

standard counterfactual analysis. To allay this worry, we need to keep in mind that weak safety does not express a *counterfactual conditional*, but rather expresses a *subjunctive conditional with a true antecedent*. Note that the hallmark of a counterfactual is that its antecedent is presumed to be *false* in the actual world “ α ”. For the sensitivity condition ($\sim p \Rightarrow \sim Bp$), the antecedent ($\sim p$) is presumed to be *false* in α because we are concerned with a class of *true beliefs* which fail to constitute knowledge, cases such as ROGER. The root idea here is that a true belief fails to constitute knowledge if it doesn’t have the property of being sensitive, such that sensitivity emerges as a non-trivial condition for knowledge, over and above the conditions that the agent must *truly believe* that p to know that p . By contrast, for weak safety ($Bp \Rightarrow p$), the presumption is that the antecedent (Bp) is *true* in α . For example, when weak safety excludes veritically lucky *true beliefs* from the extension of knowledge, the antecedent (Bp) is presumed to be true in α . Likewise, when weak safety backs-up cases of knowledge such as CHUTE, the presumption is that the antecedent (Bp) is *true* in α : because the agent has a *true belief* which constitutes knowledge. The point is that when weak safety eliminates veritic luck, and corroborates knowledge, the presumption is that the antecedent (Bp) is true in α .

We should therefore construe sensitivity as a *counterfactual subjunctive conditional* and weak safety as a *non-counterfactual subjunctive conditional*.²⁶ As we saw in Sect. 1, weak safety requires the non-standard translation key to avoid triviality, however adopting that key for weak safety does not relinquish the standard counterfactual analysis *for counterfactuals*. Providing that we keep in mind that the non-standard key is used by weak safety to unearth the truth conditions of a non-counterfactual subjunctive, one can endorse weak safety alongside the standard analysis of counterfactuals.

In summary, the real contest is not between *weak sensitivity* and *weak safety* but is instead between the full-blooded strong sensitivity condition (SS) that succumbs to CHUTE, and a weak safety condition which overcomes that counterexample. In this contest at least, weak safety can be differentially supported as the correct necessary condition for knowledge.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

²⁶ Subjunctive conditionals with true antecedents are sometimes called “counterfactual conditionals”, and safety is sometimes described as a counterfactual condition in the literature. However, by calling safety “a non-counterfactual subjunctive” I am merely saying that its antecedent, for the cases which concern us, is presumed to be true in α . Lewis (1973a, p. 3) calls such counterfactuals with true antecedents “counterfactuals that are not counterfactual”. DeRose (2004) aptly describes the safety subjunctive as a “true/true subjunctive”; this is perhaps the clearest statement of the safety condition, since insofar as truth is a condition for knowledge, the presumption is that the consequent (p) as well as the antecedent (Bp) is true in α .

References

- Becker, K. (2007). *Epistemology modalized*. London: Routledge.
- DeRose, K. (2004). Sosa, safety, sensitivity, and skeptical hypotheses. In J. Greco (Ed.), *Ernest Sosa: And his critics* (pp. 22–41). Oxford: Blackwell.
- Engel, M. (1992). Is epistemic luck compatible with knowledge? *The Southern Journal of Philosophy*, 30(2), 59–72.
- Greco, J. (2003). Virtue and luck, epistemic and otherwise. *Metaphilosophy*, 34(3), 353–366.
- Greco, J. (2012). Better safe than sensitive. In K. Becker & T. Black (Eds.), *The sensitivity principle in epistemology* (pp. 193–206). Cambridge: Cambridge University Press.
- Lewis, D. (1973a). *Counterfactuals*. Oxford: Oxford University Press.
- Lewis, D. (1973b). Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2(4), 418–446.
- Moore, G. E. (1918). The presidential address: Some judgments of perception. *Proceedings of the Aristotelian Society*, 19, 1–29.
- Moore, G. E. (1925). A defence of common sense. In J. H. Muirhead (Ed.), *Contemporary British Philosophy (second series)* (pp. 192–233). London: Allen and Unwin.
- Moore, G. E. (1939). Proof of an external world. *Proceedings of the British Academy*, 25, 273–300.
- Nozick, R. (1981). *Philosophical explanations*. Oxford: Oxford University Press.
- Pritchard, D. H. (2008). Sensitivity, safety, and anti-luck epistemology. In J. Greco (Ed.), *The Oxford handbook of scepticism* (pp. 438–453). Oxford: Oxford University Press.
- Pritchard, D. H. (2012). In defence of modest anti-luck epistemology. In K. Becker & T. Black (Eds.), *The sensitivity principle in epistemology* (pp. 173–192). Cambridge: Cambridge University Press.
- Pritchard, D. H. (2015). Anti-luck epistemology and the Gettier problem. *Philosophical Studies*, 172(1), 93–111.
- Russell, B. (1948). *Human knowledge: Its scope and limits*. London: Allen and Unwin.
- Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, 33(13), 141–153.
- Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Oxford: Oxford University Press.
- Unger, P. (1968). An analysis of factual knowledge. *The Journal of Philosophy*, 65(6), 157–170.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly*, 44(174), 65–73.
- Zalabardo, J. (2017). Safety, sensitivity and differential support. *Synthese*. <https://doi.org/10.1007/s11229-017-1645-z>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.